



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/715,066	11/17/2003	Timothy O'Brien	022438.45514	6392

44163 7590 01/10/2008
CYR & ASSOCIATES, P.A.
605 U.S. Highway 169
Suite 300
Plymouth, MN 55441

EXAMINER

REDDIG, PETER J

ART UNIT	PAPER NUMBER
----------	--------------

1642

MAIL DATE	DELIVERY MODE
-----------	---------------

01/10/2008

PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/715,066

Applicant(s)

O'BRIEN ET AL.

Examiner

Peter J. Reddig

Art Unit

1642

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 29 October 2007.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1,2,21,22,27 and 28 is/are pending in the application.
- 4a) Of the above claim(s) 21,22,27 and 28 is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1 is/are rejected.
- 7) ☒ Claim(s) 2 is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- ☐ Notice of References Cited (PTO-892)
- ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- ☒ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date 10/29/07.
- ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____
- ☐ Notice of Informal Patent Application
- ☐ Other: _____

DETAILED ACTION

1. The Amendment filed October 29, 2007 in response to the Office Action of August 27, 2007 is acknowledged and has been entered. Previously pending claims 23-26 have been cancelled, claim 1 has been amended and new claims 27 and 28 have been added.
2. Claims 1, 2, 21, 22, 27, and 28 are pending.
3. Claims 21 and 22 have been previously withdrawn from further consideration by the examiner under 37 CFR 1.142(b) as being drawn to non-elected inventions.
4. Newly submitted claims 27 and 28 are directed to an invention that is independent or distinct from the invention originally examined for the following reasons: The invention was examined as drawn to SEQ ID NO: 4 as the elected species of isolated nucleotide encoding CA125 (SEQ ID NO: 5).

Since applicant has received an action on the merits for the originally presented invention, this invention has been constructively elected by original presentation for prosecution on the merits. Accordingly, claims 27 and 28 are withdrawn from consideration as being directed to a non-elected invention. See 37 CFR 1.142(b) and MPEP § 821.03.

5. It is noted that the amendment filed October 29, 2007 is improper because claims 21 and 22 are listed as previously presented although the claims were withdrawn by the Examiner as set forth above. Although the amendment is in fact non-responsive, in the interests of compact prosecution, the invention of Claims 1 and 2, as drawn to SEQ ID NO: 4, will be examined. However, appropriate correction is required.

6. The following rejections are being maintained:

Claim Rejections - 35 USC § 102

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

7. Claim 1 remains rejected under 35 USC 102 (b) for the reasons previously set forth in the Office Action of August 27, 2007, section 8.

Applicants argue that the reference the Examiner cites is Yin and Lloyd (*J. Biol. Chem.*, July 20, 2001, 276:27371-27375). This states that the authors isolated a 5797-base pair sequence containing a stop codon but no clear 5' initiation sequence (abstract). And it is dated July 20, 2001. The alignment the Examiner shows is with Genbank locus AF361486, which is 21,112 bp (not 5797 bp) and states that it was updated on Sept. 8, 2003.

Applicants argue that the present application claims priority to U.S. provisional patent application 60/427,045, filed November 15, 2002, before the update of Genbank locus AF361486. U.S. provisional patent application 60/427,045 discloses all of SEQ ID NO: 5. Table 21 of U.S. provisional application no. 60/427,045 discloses SEQ ID NO: 162, which is the sequence of CA125 from residue 10,432 to residue 22,152 of SEQ ID NO: 5 of the present application. And Table 25 of U.S. provisional application no. 60/427,045 provides SEQ ID NO: 310, which is disclosed to be the amino terminal extension of CA125, residues 1-10,431 of SEQ ID NO: 5 of the present application. Table 30 of U.S. provisional application no. 60/427,045 discloses the 66,764-nt cDNA matching SEQ ID NO: 4 of the present application and encoding all of SEQ ID NO: 5. Accordingly, there is support for all of SEQ ID NO: 5 in the present application before the publication date of Genbank locus AF361486.

Applicants' arguments have been considered, but have not been found persuasive because the date of the rejection is based on the publication date of Yin and Lloyd, which is greater than 1 year from the priority date of November 15, 2002 of claim 1, as previously set forth and as argued by Applicants. The priority date of the specific fragments are not pertinent to claim 1 as it is drawn to SEQ ID NO: 5. Although the later, updated version of AF361486 is longer than that originally published by Yin and

Lloyd, the originally published molecule in the *J. Biol. Chem.* encodes fragments of CA125 as argued by Applicants, see below, and thus reads on the claimed invention which encompasses nucleic acid molecules that encode fragments of CA125 (SEQ ID NO: 5). Additionally, even if the rejection were to be based on AF361486 itself, which was submitted by Yin and Lloyd to GenBank (see p. 27,371 in the footnotes), AF361486 first appeared in GenBank on July 20, 2001 (see Appendix 1) and thus was published greater than 1 year from the priority date of claim 1.

Applicants argue that the Yin and Lloyd *J. Biol. Chem.* paper does not disclose the sequence of the nucleic acid isolated. It states that the nucleic acid sequence they found produced a "deduced amino acid sequence of 1890 amino acids (Fig. 3)" (page 27372 second column) and it shows the deduced amino acid sequence in Fig. 3. Alignment of residues 1-100 of the sequence shown in the top portion of Fig. 3 of Yin and Lloyd with the present SEQ ID NO: 5 shows imperfect homology with several sequences in the multiple repeat region from residues 12,070 to 21,868 of SEQ ID NO: 5. The best homology begins with residue 13721 of SEQ ID NO: 5.

Applicants argue that alignment of the sequence beginning with FNFWSS in the middle portion of Fig. 3 with SEQ ID NO:5 produced imperfect homology also with several segments of the multiple repeat domain of SEQ ID NO:5 between residues 12,070 and 21,868 of SEQ ID NO:5. The best homology begins at residue 15,004 of SEQ ID NO: 5.

Applicants argue that Alignment of the last line of sequence in Fig. 3, beginning with VLVDGYSPN with SEQ ID NO: 5 produced alignment beginning at residues 22,076 of SEQ ID NO: 5, in the carboxy terminal domain.

Applicants argue that thus, Yin and Lloyd does not disclose the actual sequence of nucleic acids that the authors discovered. The paper discloses that the nucleic acids encoded the protein sequence

shown in Fig. 3. This protein sequence is homologous with segments of the multiple repeat domain and carboxy terminal domain of CA125 (SEQ ID NO: 5), which run from amino acid residues 12,070 to 22,152 of SEQ ID NO: 5. No homology of the protein sequence disclosed in Fig. 3 of Yin and Lloyd is found with residues 1-10,431 of SEQ ID NO:5.

Applicants' arguments have been considered, but have not been found persuasive because the claimed invention encompasses nucleic acid molecules that encode fragments of CA125 (SEQ ID NO: 5) and Applicants argue that the protein sequence of Yin and Lloyd is homologous with segments of the multiple repeat domain and carboxy terminal domain of CA125 (SEQ ID NO: 5), which run from amino acid residues 12,070 to 22,152 of SEQ ID NO: 5. Specifically in response Applicants' arguments which appear to imply at times that the homology of the protein encoded by the nucleic acid molecule of Yin and Lloyd is not high compared to CA125 (SEQ ID NO: 5), an alignment of CA125 and the protein encoded by the nucleic acid of Yin and Lloyd, AF361486, is presented in Appendix 2, which shows a 97% sequence similarity between the two proteins with long stretches of identical amino acids.

Applicant's arguments have not been found persuasive and the rejection is maintained.

New Grounds of Rejection

8. Claim 2 is objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.
9. All other objections and rejections recited in the Office Action of August 27, 2007 are withdrawn.
10. No claims allowed.
11. This action is a **final rejection** and is intended to close the prosecution of this application. Applicant's reply under 37 CFR 1.113 to this action is limited either to an appeal to the Board of Patent Appeals and Interferences or to an amendment complying with the requirements set forth below.

If applicant should desire to appeal any rejection made by the examiner, a Notice of Appeal must be filed within the period for reply identifying the rejected claim or claims appealed. The Notice of Appeal must be accompanied by the required appeal fee.

If applicant should desire to file an amendment, entry of a proposed amendment after final rejection cannot be made as a matter of right unless it merely cancels claims or complies with a formal requirement made earlier. Amendments touching the merits of the application which otherwise might not be proper may be admitted upon a showing a good and sufficient reasons why they are necessary and why they were not presented earlier.

A reply under 37 CFR 1.113 to a final rejection must include the appeal form, or cancellation of, each rejected claim. The filing of an amendment after final rejection, whether or not it is entered, does not stop the running of the statutory period for reply to the final rejection unless the examiner holds the claims to be in condition for allowance. Accordingly, if a Notice of Appeal has not been filed properly within the period for reply, or any extension of this period obtained under either 37 CFR 1.136(a) or (b), the application will become abandoned.

12. Applicant's amendment necessitated the new grounds of rejection. Thus, **THIS ACTION IS MADE FINAL**. Applicant is reminded of the extension of time policy as set forth in 37 C.F.R. 1.136(a).

A SHORTENED STATUTORY PERIOD FOR RESPONSE TO THIS FINAL ACTION IS SET TO EXPIRE THREE MONTHS FROM THE DATE OF THIS ACTION. IN THE EVENT A FIRST RESPONSE IS FILED WITHIN TWO MONTHS OF THE MAILING DATE OF THIS FINAL ACTION AND THE ADVISORY ACTION IS NOT MAILED UNTIL AFTER THE END OF THE THREE-MONTH SHORTENED STATUTORY PERIOD, THEN THE SHORTENED STATUTORY PERIOD WILL EXPIRE ON THE DATE THE ADVISORY ACTION IS MAILED, AND ANY EXTENSION FEE PURSUANT TO 37 C.F.R. 1.136(a) WILL BE CALCULATED FROM THE MAILING DATE OF THE ADVISORY ACTION. IN NO EVENT WILL THE STATUTORY PERIOD FOR RESPONSE EXPIRE LATER THAN SIX MONTHS FROM THE DATE OF THIS FINAL ACTION.

13. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Peter J. Reddig whose telephone number is (571) 272-9031. The examiner can normally be reached on M-F 8:30 a.m.-5:00 p.m.

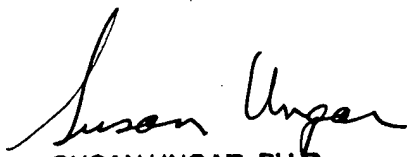
Application/Control Number:
10/715,066
Art Unit: 1642

Page 7

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Larry Helms can be reached on (571) 272-0832. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Peter J. Reddig
Examiner
Art Unit 1642



SUSAN UNGAR, PH.D
PRIMARY EXAMINER

PJR

Appendix 1.

LOCUS AF361486 5797 bp mRNA linear PRI 20-JUL-2001
DEFINITION Homo sapiens mucin 16 (MUC16) mRNA, partial cds.
ACCESSION AF361486
VERSION AF361486.1 GI:14971109
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5797)
AUTHORS Yin,B.W. and Lloyd,K.O.
TITLE Molecular cloning of the ca125 ovarian cancer antigen.
identification as a new mucin, muc16
JOURNAL J. Biol. Chem. 276 (29), 27371-27375 (2001)
PUBMED 11369781
REFERENCE 2 (bases 1 to 5797)
AUTHORS Lloyd,K.O. and Yin,B.W.T.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2001) Sloan-Kettering Institute for Cancer
Research, 1275 York Ave., New York, NY 10021, USA
COMMENT [WARNING] On Aug 26, 2003 this sequence was replaced by
gi:34223840.
FEATURES Location/Qualifiers
source 1..5797
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
gene <1..5797
/gene="MUC16"
CDS <1..5673
/gene="MUC16"
/note="ovarian cancer antigen CA125"
/codon_start=1
/product="mucin 16"
/protein_id="AAK74120.1"
/db_xref="GI:14971110"
/translation="RVDPIGPGLDRERLYWELSQLTNSITELGPYTLDRDSLYVNGFN
PWSSVPTTSTPGTSTVHLATSGTPSSLPGHTAPVPLLIPFTLNFTITNLHYEENMQHP
GSRKFNTTERVLQGLLKPLFKSTSVGPLYSGCRLTLLRPEKHGAATGVDAICTLRDLP
TGPGLDRERLYWELSQLTNSVTELGPYTLDRDSLYVNGFTHRSSVPTTSIPGTSVH
ETSGTPASLPGHTAPGPLLVPFTLNFTITNLQYEEAMRHPGSRKFNTTERVLQGLLK
LKFSTSVGPLYSGCRLTLLRPEKRGAAATGVDTICTHRLDPLNPGLDREQLYWELSKLT
RGIIELGPYLLDRGSLYVNGFTHRNFPITSTPGTSTVHLGTSETPSSLPRPIVPGPL
LVPFTLNFTITNLQYEEAMRHPGSRKFNTTERVLQGLLRPLFKNTSIGPLYSSCRLTL
LRPEKDKAATRVDAICTHHPDPQSPGLNREQLYWELSQLTGITELGPYTLDRDSLYV
DGFTHWSPIPTTSTPGTSIVNLGTSGIPPSLPETTATGPLLVPFTLNFTITNLQYEEEN

MGHPGSRKFNITESVLQGLLKPLFKSTSVGPLYSGCRLTLLRPEKDGVA TRVDAICTH
RDPDKIPGLDRQQLYWELSQLTHSITELGPYTLDRDSL YVNGFTQRSSVPTTSTPGTF
TVQPETSETPSSLPGPTATGPVLLPFTLNFTIINLQYEDMHRPGSRKFNTTERVLQ
LLMPLFKNTSVSSLYSGCRLTLLRPEKDGAATRVDVCTHRPDKSPGLDRERLYWKL
SQLTHGITE LGPYTLDRHSLYVNGFTHQSSMTTTRTPDTSTMHLATSRTPASLSGPTT
ASPLLVLFTINFTITNLRYEENMHHPGSRKFNTTERVLQGLLRPVFKNTSVGPLYSGC
RLTLLRPPKDGAA TKVDAICTYRDPDKSPGLDREQLYWELSQLTHSITELGPYTLDRD
SLYVNGFTQRSSVPTT SIPGTPVDLGTSGTPVSKPGPSAASPLLVLFTLNFTITNL
YEENMQHPGSRKFNTTERVLQGLLRSLFKSTSVGPLYSGCRLTLLRPEKDGATG VDA
ICTHHPDKSPRLDREQLYWELSQLTHNITELGPYALDNDSLFVNGFTHRSSVSTTST
PGTPTVYLGASKTPASIFGPSAASHLLILFTLNFTITNLRYEENMWP GSRKFNTTERV
LQGLLRPLFKNTSVGPLYSGCRLTLLRPEKDGEATGVDAICTHRPDTPGGLDREQLY
LELSQLTHSITELGPYTLDRDSL YVNGFTHRSSVPTTSTGVVSEEPFTLNFTINNLRY
MADMGQPGSLKFNITDNVMQHLLSPLFQRSSLGARYTGCRVIALRSVKNGAETRVDLL
CTYLQPLSGPGLPIKQVFHEL SQOTHGITRLGPYSLDKDSL YLNGYNEPGPDEPPTTP
KPATTFPLPPLSEATTAMGYHLKTLTLNFTISNLQYSPDMGKGSATFNSTEGVLQHLR
PLFQKSSMGPFFYLGCQLISLRPEKDGAATGVDTTCTYHPDPVGPGLDIQQLYWELSQL
THGVTQLGFYVLD RDSL FINGYAPQNL SIRGEYQINFHIVNWNLSNPDP TSS EYITLL
RDIQDKVTTLYKGSQ LHDTRFCLVTNLTMDSVLVTVKALFSSNLDP SLVEQVFLDKT
LNASFWHLGSTYQLVDIHVTEMES SVYQPTSSSSTQHFPNFTITNL PYSQDKAQPGT
TNYQRNKRNIEDALNQLFRNSSIKSYFSDCQVSTFRSVPNRHHTGVDSL CNFSP LARR
VDRVAIYE EFLRMTRNGTQLQNFTLDRSSVLVDGYSPNRNEPLTGNSDLPFWAVILIG
LAGLLGLITCLICGVLVTTTRRRKKEGEYNVQQQCPGYYS HLDLEDLQ"

ORIGIN

1	cgcgttgatc	ccatcggacc	tggactggac	agagagcggc	tatactggga	gctgagccag
61	ctgaccaaca	gcacacaga	gctgggaccc	tacaccctgg	atagggacag	tctctatgtc
121	aatggcttca	acccttggag	ctctgtgcca	accaccagca	ctcctgggac	ctccacagt
181	cacctggcaa	cctctgggac	tccatcctcc	ctgcctggcc	acacagcccc	tgccctctc
241	ttgataccat	tcaccctcaa	ctttaccatc	accaacctgc	attatgaaga	aaacatgcaa
301	cacctgggtt	ccaggaagtt	caacaccacg	gagaggggtc	tgcaggggtc	gctcaagccc
361	ttgttcaaga	gcaccagcgt	tggccctctg	tactctgggt	gcagactgac	cttgctcaga
421	cctgagaaac	atggggcagc	cactggagtg	gacgccatct	gcaccctccg	ccttgatccc
481	actggtcctg	gactggacag	agagcggcta	tactgggagc	tgagccagct	gaccaacagc
541	gttacagagc	tgggccccta	caccttggac	agggacagtc	tctatgtcaa	tggttccacc
601	catcggagct	ctgtgccaac	caccagtatt	cctgggacct	ctgcagtgca	cctggaaacc
661	tctgggactc	cagcctccct	ccctggccac	acagccccctg	gccctctcct	ggtgccattc
721	accctcaact	tcactatcac	caacctgcag	tatgaggagg	acatgcgtca	ccctggttcc
781	aggaagttca	acaccacgga	gagagtcctg	cagggtctgc	tcaagccctt	gttcaagagc
841	accagtgttg	gccctctgta	ctctggctgc	agactgacct	tgctcaggcc	tgaaaaacgt
901	ggggcagcca	ccggcgtgga	caccatctgc	actcaccgcc	ttgaccctct	aaacctgga
961	ctggacagag	agcagctata	ctgggagctg	agcaaactga	cccgtggcat	catcgagctg
1021	ggcccctacc	tcctggacag	aggcagtctc	tatgtcaatg	gtttcaccce	tcggaaacttt
1081	gtgcccata	ccagcactcc	tgggacctcc	acagtacacc	taggaacctc	tgaaactcca
1141	tcctccctac	ctagacccat	agtgcctggc	cctctcctgg	tgccattcac	cctcaacttc
1201	accatcacca	acttgcahta	tgaggaggcc	atgcgacacc	ctggctccag	gaagttcaat
1261	accacggaga	gggtcctaca	gggtctgctc	aggcccttgt	tcaagaatac	cagtatcggc
1321	cctctgtact	ccagctgcag	actgaccttg	ctcaggccag	agaaggacaa	ggcagccacc
1381	agagtggatg	ccatctgtac	ccaccaccct	gacctcaaaa	gccctggact	gaacagagag
1441	cagctgtact	gggagctgag	ccagctgacc	cacggcatca	ctgagctggg	cccctacacc
1501	ctggacaggg	acagtctcta	tgtcgatggg	ttcactcatt	ggagccccc	accaaccacc
1561	agcactcctg	ggacctccat	agtgaacctg	ggaacctctg	ggatccccc	ttccctccct
1621	gaaactacag	ccaccggccc	tctcctgggt	ccattcacac	tcaacttcac	catcactaac
1681	ctacagtatg	aggagaacat	gggtcaccct	ggctccagga	agttcaacat	cacggagagt

1741	gttctgcagg	gtctgctcaa	gcccttggtc	aagagcacca	gtgttgcccc	tctgtattct
1801	ggctgcagac	tgaccttgct	caggcctgag	aaggacggag	tagccaccag	agtggacgcc
1861	atctgcaccc	accgccctga	ccccaaaatc	cctgggctag	acagacagca	gctatactgg
1921	gagctgagcc	agctgaccca	cagcatcact	gagctgggac	cctacaccct	ggatagggac
1981	agtctctatg	tcaatggttt	caccacggc	agctctgtgc	ccaccaccag	cactcctggg
2041	actttcacag	tacagccgga	aacctctgag	actccatcat	ccctccctgg	ccccacagcc
2101	actggccctg	tctgtctgcc	attcaccctc	aattttacca	tcattaacct	gcagtatgag
2161	gaggacatgc	atcgccctgg	ctccaggaag	ttcaacacca	cggagagggg	ccttcagggg
2221	ctgcttatgc	ccttggtcaa	gaacaccagt	gtcagctctc	tgtactctgg	ttgcagactg
2281	accttgctca	ggcctgagaa	ggatggggca	gccaccagag	tggatgctgt	ctgcacccat
2341	cgccctgacc	ccaaaagccc	tggactggac	agagagcggc	tgtactggaa	gctgagccag
2401	ctgaccacag	gcatcactga	gctgggcccc	tacaccctgg	acaggcacag	tctctatgtc
2461	aatggtttca	cccatcagag	ctctatgacg	accaccagaa	ctcctgatac	ctccacaatg
2521	cacctggcaa	cctcgagaac	tccagcctcc	ctgtctggac	ctacgaccgc	cagccctctc
2581	ctggtgctat	tcacaattaa	cttcaccatc	actaacctgc	ggtatgagga	gaacatgcat
2641	cacctgggct	ctagaaagtt	taacaccacg	gagagagtcc	ttcagggctc	gctcaggcct
2701	gtgttcaaga	acaccagtgt	tggccctctg	tactctggct	gcagactgac	cttgtcagg
2761	cccaagaagg	atggggcagc	caccaaaagt	gatgccatct	gcacctaccg	ccctgatccc
2821	aaaagccctg	gactggacag	agagcagcta	tactgggagc	tgagccagct	aaccacagc
2881	atcactgagc	tgggccccta	caccttgga	agggacagtc	tctatgtcaa	tggtttcaca
2941	cagcggagct	ctgtgcccac	cactagcatt	cctgggaccc	ccacagtgga	cctgggaaca
3001	tctgggactc	cagttttctaa	acctgggtccc	tcggctgcca	gccctctcct	ggtgctattc
3061	actctcaact	tcaccatcac	caacctgcgg	tatgaggaga	acatgcagca	ccctgggtcc
3121	aggaagtcca	acaccacgga	gagggtcctt	cagggcctgc	tcagggtccct	gttcaagagc
3181	accagtgttg	gccctctgta	ctctggctgc	agactgactt	tgctcaggcc	tgaaaaggat
3241	gggacagcca	ctggagtgga	tgccatctgc	accaccacc	ctgaccccaa	aagccctagg
3301	ctggacagag	agcagctgta	ttgggagctg	agccagctga	cccacaatat	cactgagctg
3361	ggcccttatg	ccctggacaa	cgacagcctc	tttgtcaatg	gtttcactca	tcggagctct
3421	gtgtccacca	ccagcactcc	tgggaccccc	acagtgtatc	tgggagcatc	taagactcca
3481	gcctcgatat	ttggcccttc	agctgccagc	catctcctga	tactattcac	cctcaacttc
3541	accatcacta	acctgcggta	tgaggagaac	atgtggcctg	gctccaggaa	gttcaacact
3601	acagagaggg	tccttcaggg	cctgctaagg	cccttggtca	agaacaccag	tgttgccct
3661	ctgtactctg	gctgcaggct	gaccttgctc	aggccagaga	aagatgggga	agccaccgga
3721	gtggatgcca	tctgcaccca	ccgccctgac	cccacaggcc	ctgggctgga	cagagagcag
3781	ctgtatttgg	agctgagcca	gctgaccac	agcatcactg	agctgggccc	ctacacactg
3841	gacagggaca	gtctctatgt	caatggtttc	acccatcgga	gctctgtacc	caccaccagc
3901	accgggggtg	tcagcgagga	gccattcaca	ctgaacttca	ccatcaacaa	cctgcgtac
3961	atggcggaca	tgggccaacc	cggctccctc	aagttcaaca	tcacagacaa	cgtcatgcag
4021	cacctgctca	gtcctttggt	ccagaggagc	agcctgggtg	cacggtacac	aggctgcagg
4081	gtcatcgcac	taaggctctgt	gaagaacggg	gctgagacac	gggtggacct	cctctgcacc
4141	tacctgcagc	ccctcagcgg	cccaggctctg	cctatcaagc	aggtgttcca	tgagctgagc
4201	cagcagaccc	atggcatcac	ccggctgggc	ccctactctc	tggacaaaaga	cagcctctac
4261	cttaacgggt	acaatgaacc	tgggtccagat	gagcctccta	caactcccaa	gccagccacc
4321	acattcctgc	ctcctctgtc	agaagccaca	acagccatgg	ggtaccacct	gaagaccctc
4381	acactcaact	tcaccatctc	caatctccag	tattcaccag	atatgggcaa	gggtcagct
4441	acattcaact	ccaccgaggg	ggctccttcag	cacctgctca	gacccttggt	ccagaagagc
4501	agcatgggcc	ccttctactt	gggttgccaa	ctgatctccc	tcaggcctga	gaaggatggg
4561	gcagccactg	gtgtggacac	cacctgcacc	taccaccctg	accctgtggg	ccccgggctg
4621	gacatacagc	agctttactg	ggagctgagt	cagctgaccc	atggtgtcac	ccaactgggc
4681	ttctatgtcc	tggacagggg	tagcctcttc	atcaatggct	atgcacccca	gaatttatca
4741	atccggggcg	agtaccagat	aaatttccac	attgtcaact	ggaacctcag	taatccagac
4801	cccacatcct	cagagtacat	cacctgctg	agggacatcc	aggacaaggt	caccacactc
4861	tacaaaggca	gtcaactaca	tgacacattc	cgcttctgcc	tggtcaccaa	cttgacgatg

```

4921 gactccgtgt tggteactgt caaggcattg ttctectcca atttggaccc cagcctgggtg
4981 gagcaagtct ttctagataa gaccctgaat gcctcattcc attggctggg ctccacctac
5041 cagttgggtg acatccatgt gacagaaatg gagtcacag tttatcaacc aacaagcagc
5101 tccagcacc agcacttcta cccgaatttc accatcacca acctaccata tcccaggac
5161 aaagcccagc caggcaccac caattaccag aggaacaaaa ggaatattga ggatgcgctc
5221 aaccaactct tccgaaacag cagcatcaag agttatTTTT ctgactgtca agtttcaaca
5281 ttcaggtctg tccccaacag gcaccacacc ggggtggact ccctgtgtaa cttctcgcca
5341 ctggctcgga gagtagacag agttgccatc tatgaggaat ttctgcggat gacccggaat
5401 ggtaccacag tgcagaactt caccctggac aggagcagtg tccttgtgga tgggtattct
5461 cccaacagaa atgagccctt aactgggaat tctgacctc ccttctgggc tgtcatcctc
5521 atcggtctgg caggactcct gggactcatc acatgcctga tctgcgggtg cctggtgacc
5581 accgcgcggc ggaagaagga aggagaatac aacgtccagc aacagtgcc aggctactac
5641 cagtcacacc tagacctgga ggatctgcaa tgactggaac ttgccgggtg ctgggggtgc
5701 tttccccccag ccagggtcca aagaagcttg gctggggcag aaataaacca tattggtcgg
5761 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa

```

Appendix 2

> sp|Q8WXI7.2|MUC16_HUMAN **GD** Mucin-16 (Ovarian carcinoma antigen CA125) (Ovarian cancer-related tumor marker CA125) (CA-125)

Score = 3693 bits (9577), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 1845/1890 (97%), Positives = 1846/1890 (97%), Gaps = 0/1890 (0%)

Query	1	RVDPIGPGDLRERLYWELSQLTNSITELGPYTLDRDSL VNGFNPWSSVPTTSTPGTSTV	60
		R DP PGL E LYWELS LT I ELGPYTLDR SLYVNGF S PTTSTPGTSTV	
Sbjct	20263	RPDPKSPGLXXEXLYWELSQLTXXIXELGPYTLDRXSL VNGFTHXXSXPTTSTPGTSTV	20322
Query	61	HLATSGTPSSLPGHTAPVPLLPFTLNFTITNLHYEENMQHPGSRKFNTTTERVLQGLLKP	120
		TSGTPSS P TAPVPLLPFTLNFTITNLHYEENMQHPGSRKFNTTTERVLQGLLKP	
Sbjct	20323	XXGTSGTPSSXPXTAPVPLLPFTLNFTITNLHYEENMQHPGSRKFNTTTERVLQGLLKP	20382
Query	121	LFKSTSVGPLYSGCRLTLLRPEKHGAATGVDAICTLRDPTGPGDLRERLYWELSQLTNS	180
		LFKSTSVGPLYSGCRLTLLRPEKHGAATGVDAICTLRDPTGPGDLRERLYWELSQLTNS	
Sbjct	20383	LFKSTSVGPLYSGCRLTLLRPEKHGAATGVDAICTLRDPTGPGDLRERLYWELSQLTNS	20442
Query	181	VTELGPYTLDRDSL VNGFTHRSSVPTTSIPGTS AVHLETSGTPASLPGHTAPGPLLVPF	240
		VTELGPYTLDRDSL VNGFT RSSVPTTSIPGTS AVHLETSGTPASLPGHTAPGPLLVPF	
Sbjct	20443	VTELGPYTLDRDSL VNGFTQRSSVPTTSIPGTS AVHLETSGTPASLPGHTAPGPLLVPF	20502
Query	241	TLNFTITNLQYEEDMRHPGSRKFNTTTERVLQGLLKL PFKSTSVGPLYSGCRLTLLRPEKR	300
		TLNFTITNLQYE DMRHPGSRKFNTTTERVLQGLLKL PFKSTSVGPLYSGCRLTLLRPEKR	
Sbjct	20503	TLNFTITNLQYEVDMRHPGSRKFNTTTERVLQGLLKL PFKSTSVGPLYSGCRLTLLRPEKR	20562
Query	301	GAATGVDTICTHRLDPLNPGLDREQLYWELSKLTRGIIELGPYLLDRGSL VNGFTHRNF	360
		GAATGVDTICTHRLDPLNPGLDREQLYWELSKLTRGIIELGPYLLDRGSL VNGFTHRNF	
Sbjct	20563	GAATGVDTICTHRLDPLNPGLDREQLYWELSKLTRGIIELGPYLLDRGSL VNGFTHRNF	20622
Query	361	VPITSTPGTSTVHLGTSETPSSLPRIVPGPPLLVPFTLNFTITNLQYEEAMRHPGSRKFN	420
		VPITSTPGTSTVHLGTSETPSSLPRIVPGPPLLVPFTLNFTITNLQYEEAMRHPGSRKFN	

Sbjct	20623	VPITSTPGTSTVHLGTSETPSSLRPPIVPGPLLVPFTLNFTITNLQYEEAMRHGSRKFN	20682
Query	421	TTERVLQGLLRPLFKNTSIGPLYSSCRLTLLRPEKDKAATRVDICTHHPDPQSPGLNRE	480
		TTERVLQGLLRPLFKNTSIGPLYSSCRLTLLRPEKDKAATRVDICTHHPDPQSPGLNRE	
Sbjct	20683	TTERVLQGLLRPLFKNTSIGPLYSSCRLTLLRPEKDKAATRVDICTHHPDPQSPGLNRE	20742
Query	481	QLYWELSQLTHGITELGPYTLDRDSLVDGFTHWSPITPTSTPGTSIVNLGTSGIPPSLP	540
		QLYWELSQLTHGITELGPYTLDRDSLVDGFTHWSPITPTSTPGTSIVNLGTSGIPPSLP	
Sbjct	20743	QLYWELSQLTHGITELGPYTLDRDSLVDGFTHWSPITPTSTPGTSIVNLGTSGIPPSLP	20802
Query	541	ETTATGPLLVPFTLNFTITNLQYEEENMGHPSRKFNITESVLQGLLKPLFKSTSVGPLY	600
		ETT PLL PFT N TITNL M PGSRKFN TE VLQGLLKPLFKSTSVGPLY	
Sbjct	20803	ETTXXXPLLXPFTXNXTITNLXXXXXXMXXPGSRKFNTTTERVLQGLLKPLFKSTSVGPLY	20862
Query	601	GCRLTLLRPEKDGVA TRVDAICTHRDPKIPGLDRQQLYWELSQLTHSITELGPYTLDRD	660
		GCRLTLLRPEKDGVA TRVDAICTHRDPKIPGLDRQQLYWELSQLTHSITELGPYTLDRD	
Sbjct	20863	GCRLTLLRPEKDGVA TRVDAICTHRDPKIPGLDRQQLYWELSQLTHSITELGPYTLDRD	20922
Query	661	SLYVNGFTQRSSVPTTSTPGTFTVQPETSETPSSLPGPTATGPVLLPFTLNFTIINLQYE	720
		SLYVNGFTQRSSVPTTSTPGTFTVQPETSETPSSLPGPTATGPVLLPFTLNFTI NLQYE	
Sbjct	20923	SLYVNGFTQRSSVPTTSTPGTFTVQPETSETPSSLPGPTATGPVLLPFTLNFTITNLQYE	20982
Query	721	EDMHRPGSRKFNTTTERVLQGLLMPLFKNTSVSSLYSGCRLTLLRPEKDGAATRVDVCTH	780
		EDMHRPGSRKFNTTTERVLQGLLMPLFKNTSVSSLYSGCRLTLLRPEKDGAATRVDVCTH	
Sbjct	20983	EDMHRPGSRKFNTTTERVLQGLLMPLFKNTSVSSLYSGCRLTLLRPEKDGAATRVDVCTH	21042
Query	781	RPDPKSPGLDRERLYWKLSQLTHGITELGPYTLDRHSLYVNGFTHQSSMTTTRTPDTSTM	840
		RPDPKSPGLDRERLYWKLSQLTHGITELGPYTLDRHSLYVNGFTHQSSMTTTRTPDTSTM	
Sbjct	21043	RPDPKSPGLDRERLYWKLSQLTHGITELGPYTLDRHSLYVNGFTHQSSMTTTRTPDTSTM	21102
Query	841	HLATSRTPASLSGPTTASPLLVLFTINFTITNLRYEENMHHPGSRKFNTTTERVLQGLLRP	900
		HLATSRTPASLSGPTTASPLLVLFTINFTITNLRYEENMHHPGSRKFNTTTERVLQGLLRP	
Sbjct	21103	HLATSRTPASLSGPTTASPLLVLFTINFTITNLRYEENMHHPGSRKFNTTTERVLQGLLRP	21162
Query	901	VFKNTSVGPLYSGCRLTLLRPKKDGAATKVDAICTYRPDPKSPGLDREQLYWELSQLTHS	960
		VFKNTSVGPLYSGCRLTLLRPKKDGAATKVDAICTYRPDPKSPGLDREQLYWELSQLTHS	
Sbjct	21163	VFKNTSVGPLYSGCRLTLLRPKKDGAATKVDAICTYRPDPKSPGLDREQLYWELSQLTHS	21222
Query	961	ITELGPYTLDRDSLVDGFTQRSSVPTTSIPGTPTVDLGTSGTPVSKPGPSAASPLLVL	1020
		ITELGPYTLDRDSLVDGFTQRSSVPTTSIPGTPTVDLGTSGTPVSKPGPSAASPLLVL	
Sbjct	21223	ITELGPYTLDRDSLVDGFTQRSSVPTTSIPGTPTVDLGTSGTPVSKPGPSAASPLLVL	21282
Query	1021	TLNFTITNLRYEENMQHPGSRKFNTTTERVLQGLLRSLFKSTSVGPLYSGCRLTLLRPEKD	1080
		TLNFTITNLRYEENMQHPGSRKFNTTTERVLQGLLRSLFKSTSVGPLYSGCRLTLLRPEKD	
Sbjct	21283	TLNFTITNLRYEENMQHPGSRKFNTTTERVLQGLLRSLFKSTSVGPLYSGCRLTLLRPEKD	21342
Query	1081	GTATGVDAICTHHPDPKSPRLDREQLYWELSQLTHNITELGPYALDNDLSLVNGFTHR	1140
		GTATGVDAICTHHPDPKSPRLDREQLYWELSQLTHNITELG YALDNDLSLVNGFTHR	
Sbjct	21343	GTATGVDAICTHHPDPKSPRLDREQLYWELSQLTHNITELGHYALDNDLSLVNGFTHR	21402
Query	1141	VSTTSTPGTPTVYLGAASKTPASIFGPSAASHLLILFTLNFTITNLRYEENMWPGSRKF	1200
		VSTTSTPGTPTVYLGAASKTPASIFGPSAASHLLILFTLNFTITNLRYEENMWPGSRKF	
Sbjct	21403	VSTTSTPGTPTVYLGAASKTPASIFGPSAASHLLILFTLNFTITNLRYEENMWPGSRKF	21462

Query	1201	TERVLQGLLRPLFKNTSVGPLYSGCRLTLLRPEKDGEATGVDAICTHRPDPTGPGLDREQ	1260
		TERVLQGLLRPLFKNTSVGPLYSG RLTLRPEKDGEATGVDAICTHRPDPTGPGLDREQ	
Sbjct	21463	TERVLQGLLRPLFKNTSVGPLYSGSRLTLLRPEKDGEATGVDAICTHRPDPTGPGLDREQ	21522
Query	1261	LYLELSQLTHSITELGPYTLDLDRSLYVNGFTHRSSVPTTSTGVVSEEPFTLNFTINNLR	1320
		LYLELSQLTHSITELGPYTLDLDRSLYVNGFTHRSSVPTTSTGVVSEEPFTLNFTINNLR	
Sbjct	21523	LYLELSQLTHSITELGPYTLDLDRSLYVNGFTHRSSVPTTSTGVVSEEPFTLNFTINNLR	21582
Query	1321	MADMGQPGSLKFNITDNVMQHLLSPLFQRSSLGARYTGCRVIALRSVKNGAETRVDLLCT	1380
		MADMGQPGSLKFNITDNVM+HLLSPLFQRSSLGARYTGCRVIALRSVKNGAETRVDLLCT	
Sbjct	21583	MADMGQPGSLKFNITDNVMKHLLSPLFQRSSLGARYTGCRVIALRSVKNGAETRVDLLCT	21642
Query	1381	YLQPLSGPGLPIKQVFHELSSQOQTHGITRLGPYSLDKDSLNLNGYNEPGPDEPPTTPKPAT	1440
		YLQPLSGPGLPIKQVFHELSSQOQTHGITRLGPYSLDKDSLNLNGYNEPG DEPPTTPKPAT	
Sbjct	21643	YLQPLSGPGLPIKQVFHELSSQOQTHGITRLGPYSLDKDSLNLNGYNEPGLDEPPTTPKPAT	21702
Query	1441	TFLPPLSEATTAMGYHLKTLTLNFTISNLQYSPDMGKSATFNSTEGVLQHLLRPLFQKS	1500
		TFLPPLSEATTAMGYHLKTLTLNFTISNLQYSPDMGKSATFNSTEGVLQHLLRPLFQKS	
Sbjct	21703	TFLPPLSEATTAMGYHLKTLTLNFTISNLQYSPDMGKSATFNSTEGVLQHLLRPLFQKS	21762
Query	1501	SMGPFYLGQCQLISLRPEKDGAATGVDTTCTYHPDPVGPGLDIQQLYWELSQLTHGVTQLG	1560
		SMGPFYLGQCQLISLRPEKDGAATGVDTTCTYHPDPVGPGLDIQQLYWELSQLTHGVTQLG	
Sbjct	21763	SMGPFYLGQCQLISLRPEKDGAATGVDTTCTYHPDPVGPGLDIQQLYWELSQLTHGVTQLG	21822
Query	1561	FYVLDRDSLFIINGYAPQNLISIRGEYQINFHIVNWNLSNPDPTSSEYITLLRDIQDKVTTL	1620
		FYVLDRDSLFIINGYAPQNLISIRGEYQINFHIVNWNLSNPDPTSSEYITLLRDIQDKVTTL	
Sbjct	21823	FYVLDRDSLFIINGYAPQNLISIRGEYQINFHIVNWNLSNPDPTSSEYITLLRDIQDKVTTL	21882
Query	1621	YKGSQHLDTFRFCLVTNLTMDSVLVTVKALFSSNLDPSLVEQVFLDKTLNASFHWLGSTY	1680
		YKGSQHLDTFRFCLVTNLTMDSVLVTVKALFSSNLDPSLVEQVFLDKTLNASFHWLGSTY	
Sbjct	21883	YKGSQHLDTFRFCLVTNLTMDSVLVTVKALFSSNLDPSLVEQVFLDKTLNASFHWLGSTY	21942
Query	1681	QLVDIHVTEMESSVYQPTSSSSTQHFPNFTITNLPYSQDKAQPGTTNYQRNKRNIEDAL	1740
		QLVDIHVTEMESSVYQPTSSSSTQHFP NFTITNLPYSQDKAQPGTTNYQRNKRNIEDAL	
Sbjct	21943	QLVDIHVTEMESSVYQPTSSSSTQHFPNFTITNLPYSQDKAQPGTTNYQRNKRNIEDAL	22002
Query	1741	NQLFRNSSIKSYFSDCQVSTFRSVPNRHHTGVDSL CNFSP LARRVDRVAIYEEFLRMTRN	1800
		NQLFRNSSIKSYFSDCQVSTFRSVPNRHHTGVDSL CNFSP LARRVDRVAIYEEFLRMTRN	
Sbjct	22003	NQLFRNSSIKSYFSDCQVSTFRSVPNRHHTGVDSL CNFSP LARRVDRVAIYEEFLRMTRN	22062
Query	1801	GTQLQNFTLDRSSVLVDGYSPNRNEPLTGNSDL PFWAVILIGLAGLLGLITCLICGVLVT	1860
		GTQLQNFTLDRSSVLVDGYSPNRNEPLTGNSDL PFWAVILIGLAGLLGLITCLICGVLVT	
Sbjct	22063	GTQLQNFTLDRSSVLVDGYSPNRNEPLTGNSDL PFWAVILIGLAGLLGLITCLICGVLVT	22122
Query	1861	TRRRKKEGEYNVQQQCPGYYQSHLDLEDLQ	1890
		TRRRKKEGEYNVQQQCPGYYQSHLDLEDLQ	
Sbjct	22123	TRRRKKEGEYNVQQQCPGYYQSHLDLEDLQ	22152